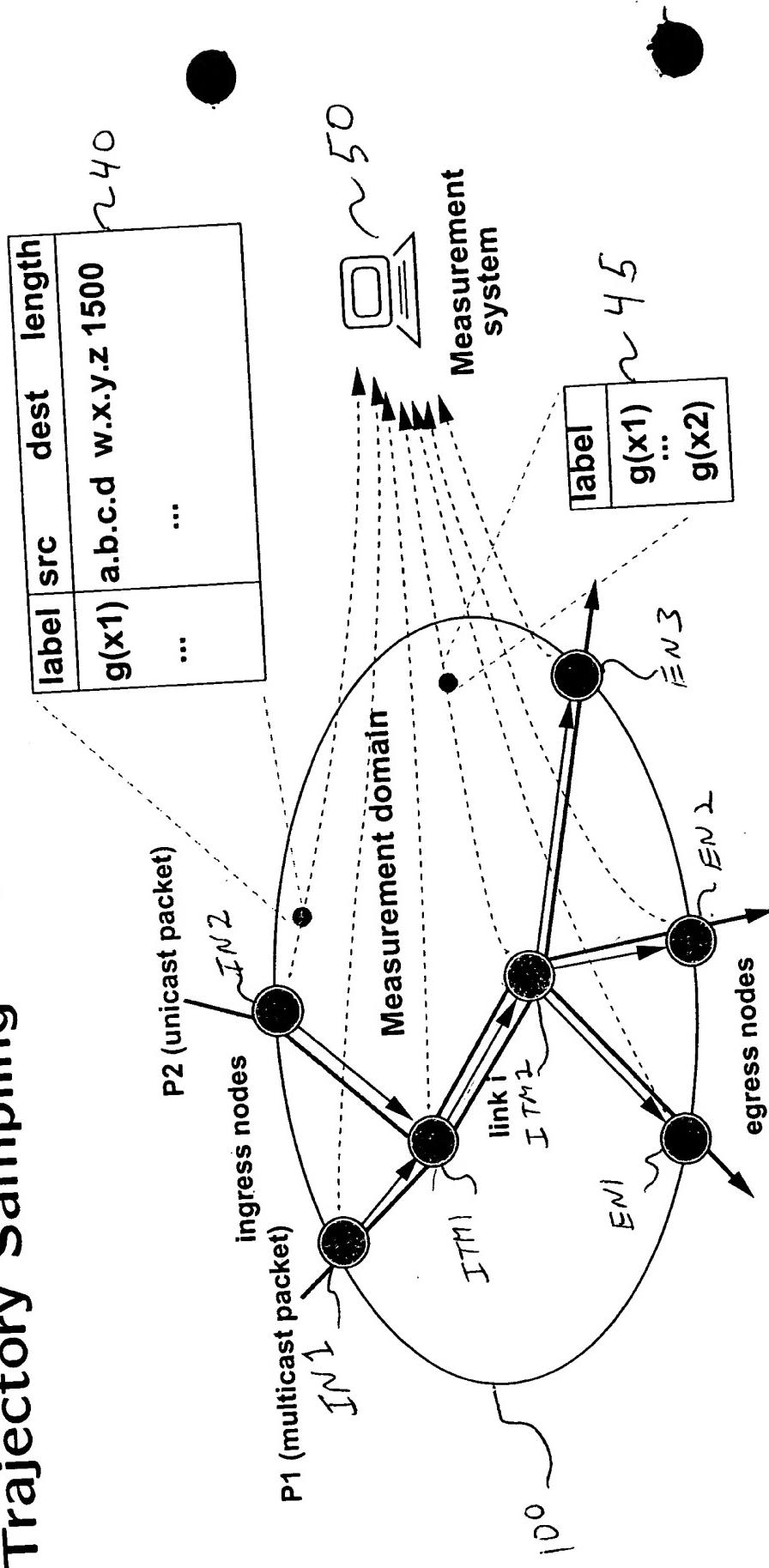


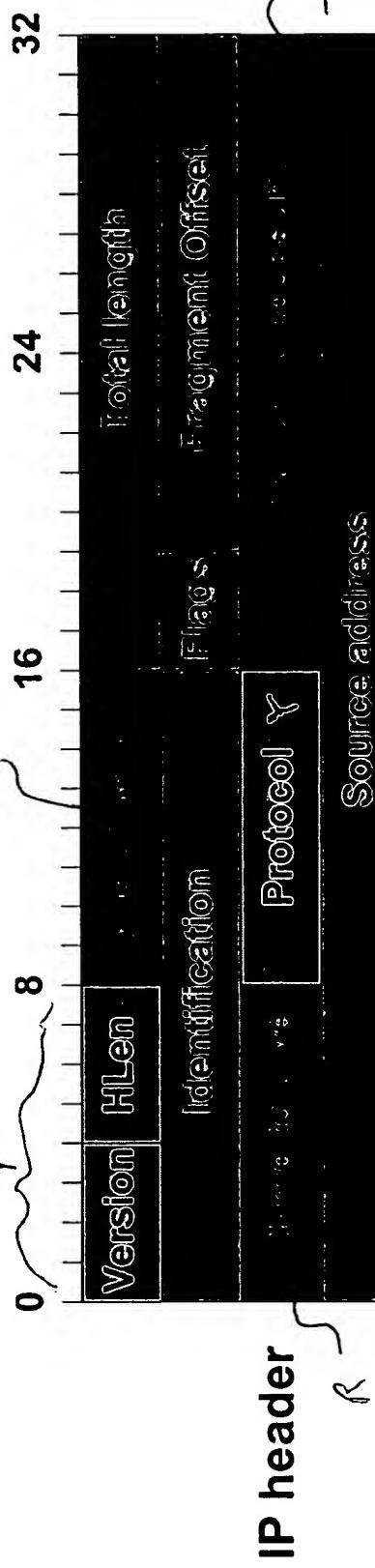
Trajectory Sampling



- Collect fields of interest only once (ingress)
- Multicast requires no special treatment

ପରିମାଣିତ କରିବାର ପାଇଁ

Fields Included in Hashes



6

Protocol (ପ୍ରୋଟୋକୋଲ୍) | ଡାକ୍ତରିକ୍ ପାଇଁ

DSCP (ଡିସଟରିଫିନିଚନ୍ ପାଇଁ) | ଶାର୍ଟ୍ ପାଇଁ

Checksum (ଚକ୍ଷମ୍ ପାଇଁ) | ଡାକ୍ତରିକ୍ ପାଇଁ

Source IP Address (ସୌରୀପ୍ୟ ପାଇଁ) | ଡାକ୍ତରିକ୍ ପାଇଁ

Destination IP Address (ଦିନ୍ଦୁପ୍ୟ ପାଇଁ) | ଡାକ୍ତରିକ୍ ପାଇଁ

Protocol (ପ୍ରୋଟୋକୋଲ୍) | ଡାକ୍ତରିକ୍ ପାଇଁ

Acknowledgment Number (ଅକ୍ଷେତ୍ର ପାଇଁ) | ଡାକ୍ତରିକ୍ ପାଇଁ

Offset (ଓଫ୍‌ସେଟ୍ ପାଇଁ) | ଡାକ୍ତରିକ୍ ପାଇଁ

Flags (ଫଲ୍ସ ପାଇଁ) | ଡାକ୍ତରିକ୍ ପାଇଁ

Window Size (ଓନ୍‌ପାଇଁ) | ଡାକ୍ତରିକ୍ ପାଇଁ

Checksum (ଚକ୍ଷମ୍ ପାଇଁ) | ଡାକ୍ତରିକ୍ ପାଇଁ

Urgent Pointer (ଉର୍ଜେଟ୍ ପାଇଁ) | ଡାକ୍ତରିକ୍ ପାଇଁ

• •

ମୁଖ୍ୟମ୍ ପାଇଁ

Low=ଅଗ୍ରତମ୍ ଯ

R R

0000000000000000

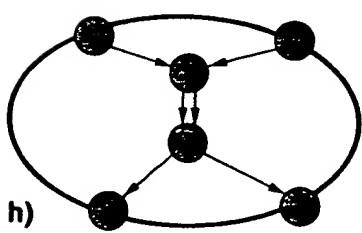
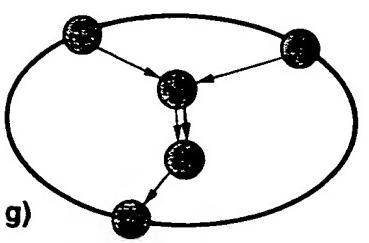
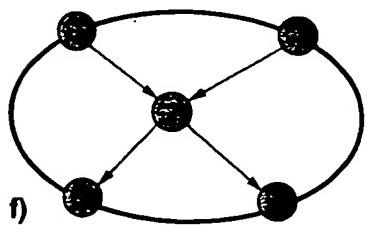
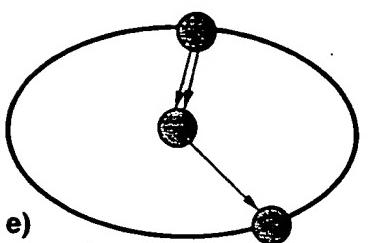
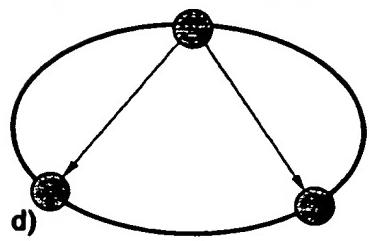
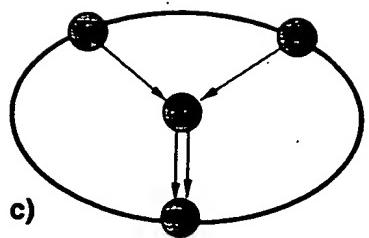
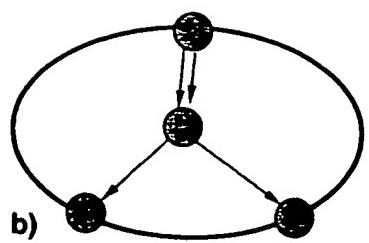
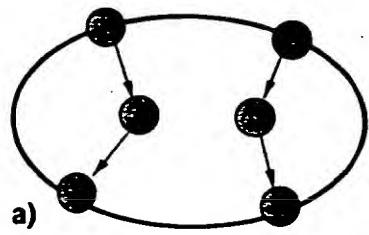
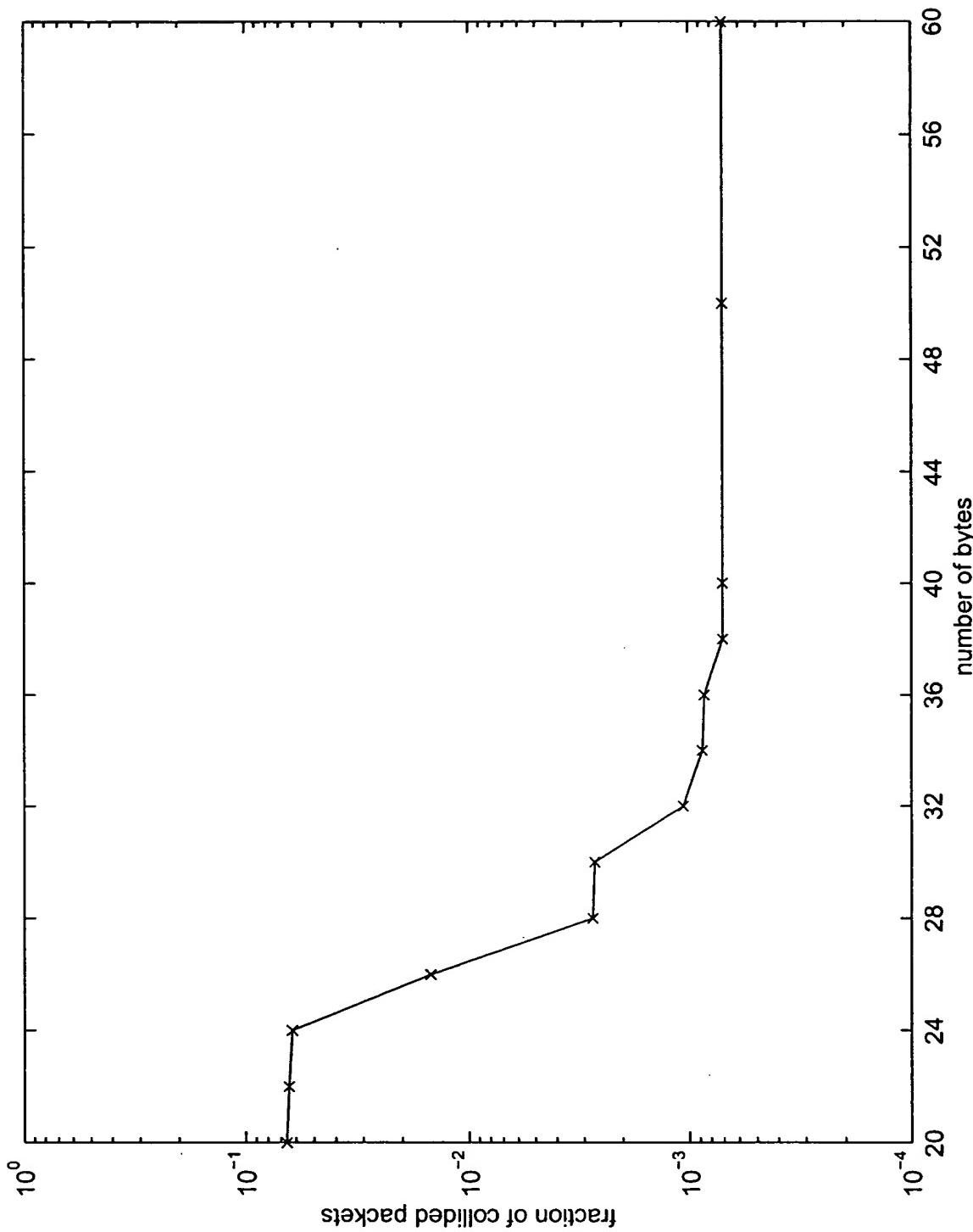


Figure 3

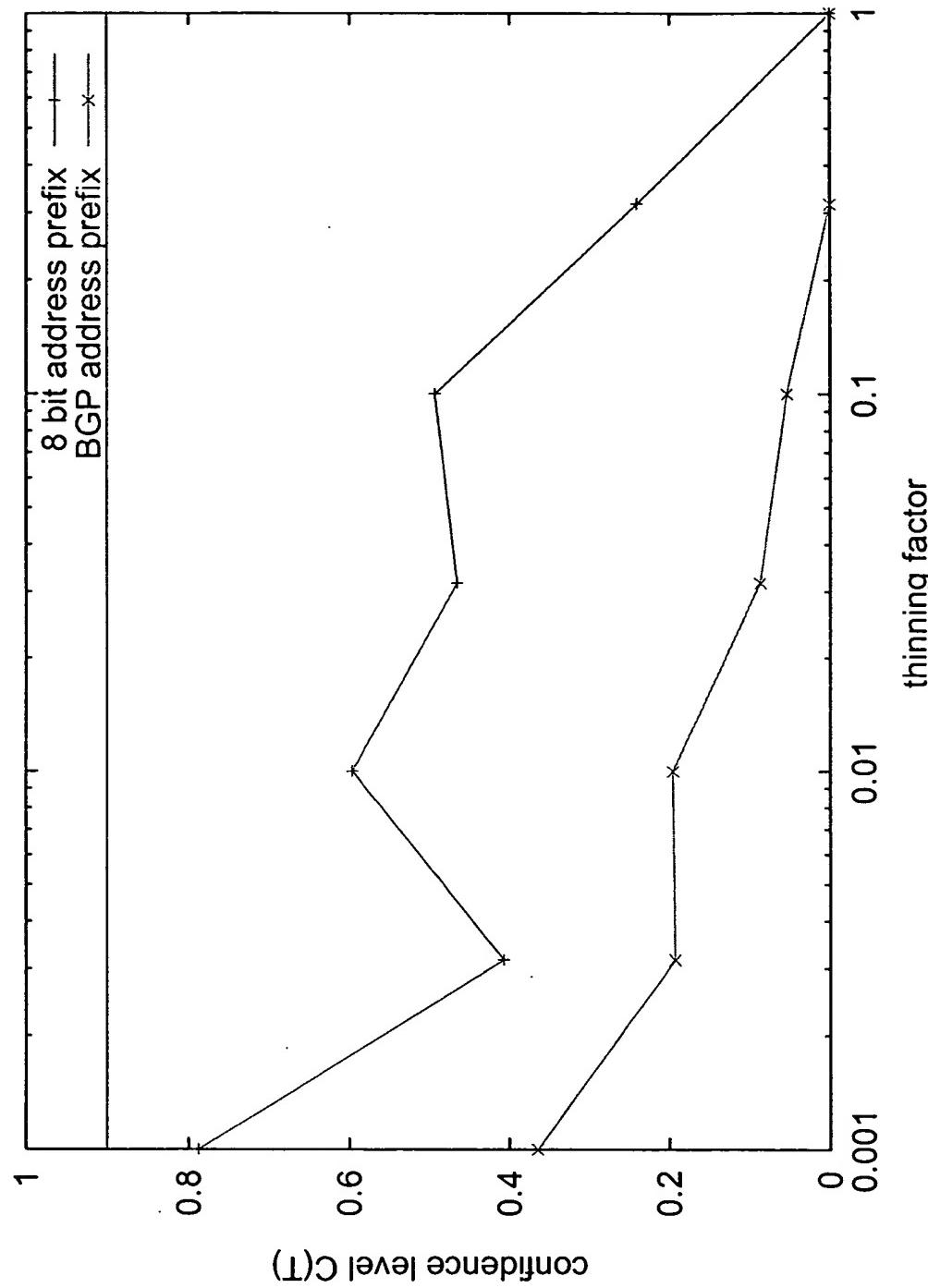
Collisions: Identical Packets are Rare



CONFIDENCE INTERVALS A

χ^2 -Test for Independence of Sampling Decision & Addresses

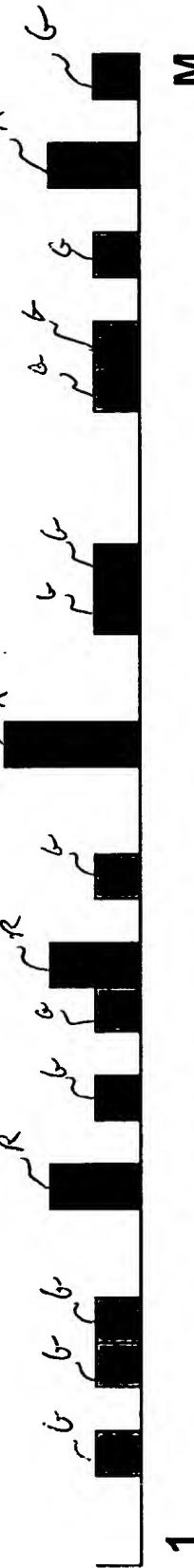
- If $C(T) < 1 - \text{significance level} \Rightarrow \text{accept hypothesis}$



Optimal Sampling

DATA FLOW LOGIC

labels collected in a measurement period



- Fix amount of measurement traffic c per period
- Tradeoff: collisions vs. label size
- Problem:

- n : number of samples in sampling period
- M : alphabet size, $m = \log_2 M$ [bits/label]
- $n \cdot m$: total amount of measurement traffic [bits]
- Goal: maximize number of *unique* labels subject to $n \cdot m \leq c$.
- Optimal alphabet size: $M^* = c \log(2)$
Optimal number of samples: $n^* = \frac{M^*}{\log(M^*)}$

Example: $c = 10^6$ bit $\Rightarrow m^* = 19.4$ bit/label
 $n^* = 5.15 \cdot 10^4$ samples

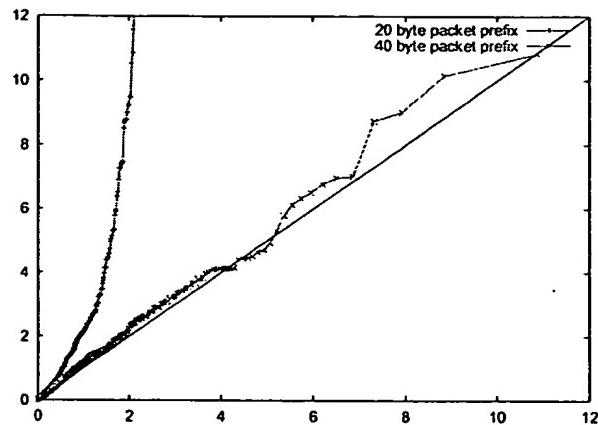


Figure 6: HASH-SAMPLED ADDRESS BITS DISTRIBUTIONS. Quantile-quantile plot of address bit chi-square values vs. chi-squared distribution with 1 degree of freedom; for various traces, primes A , thinning factors r/A ; see text. Close agreement for 40 byte packet prefixes; marked disagreement for 20 byte packet prefixes (i.e. no payload included for sampling hash)

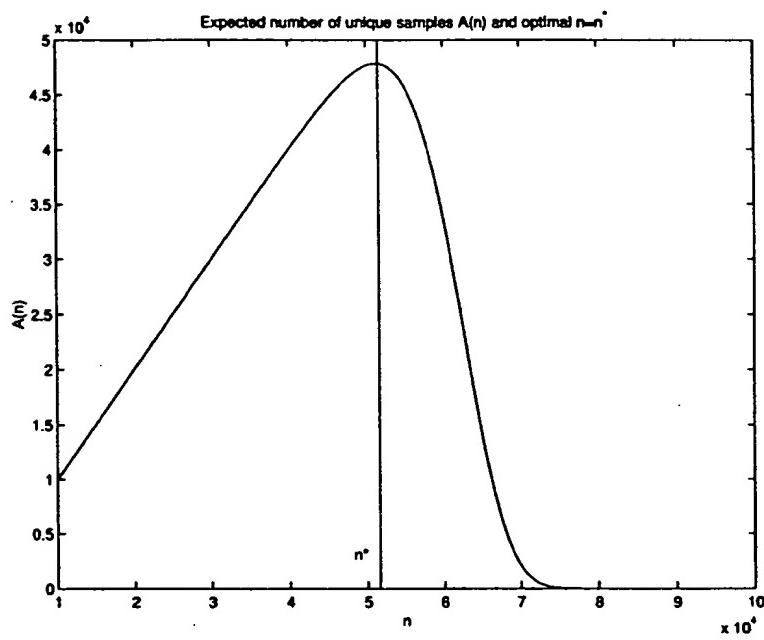
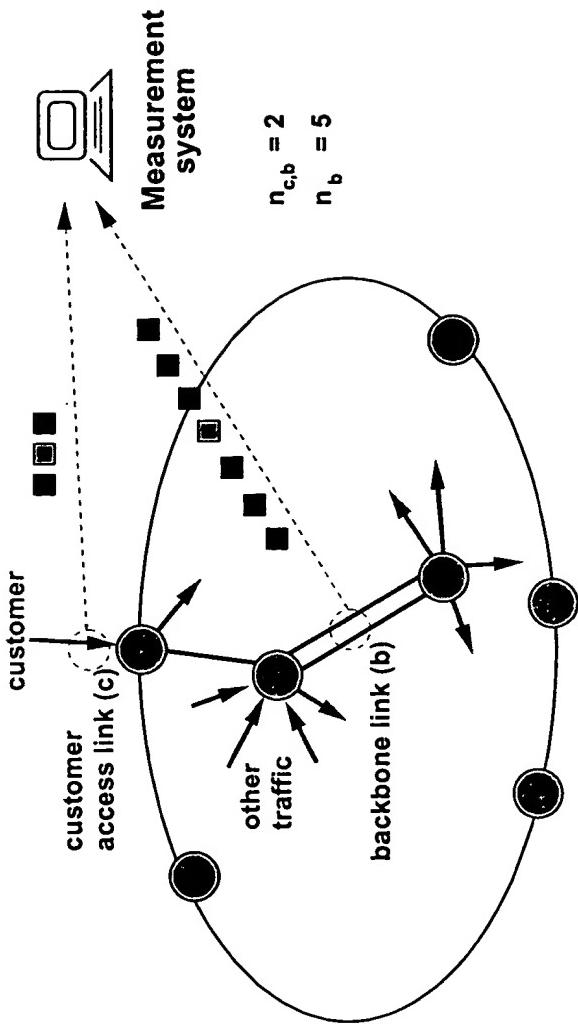


Figure 7: The expected number of unique samples $A(n)$ as a function of n , for $c = 10^6$ bit. The optimal number of samples n^* is approximately $5.15 \cdot 10^4$, with $m^* = 19.4$ bit per label. The collision probability p_{coll} is approximately 0.072, i.e., 7.2% of the samples transmitted to the collection system have to be discarded.

Inference Experiment

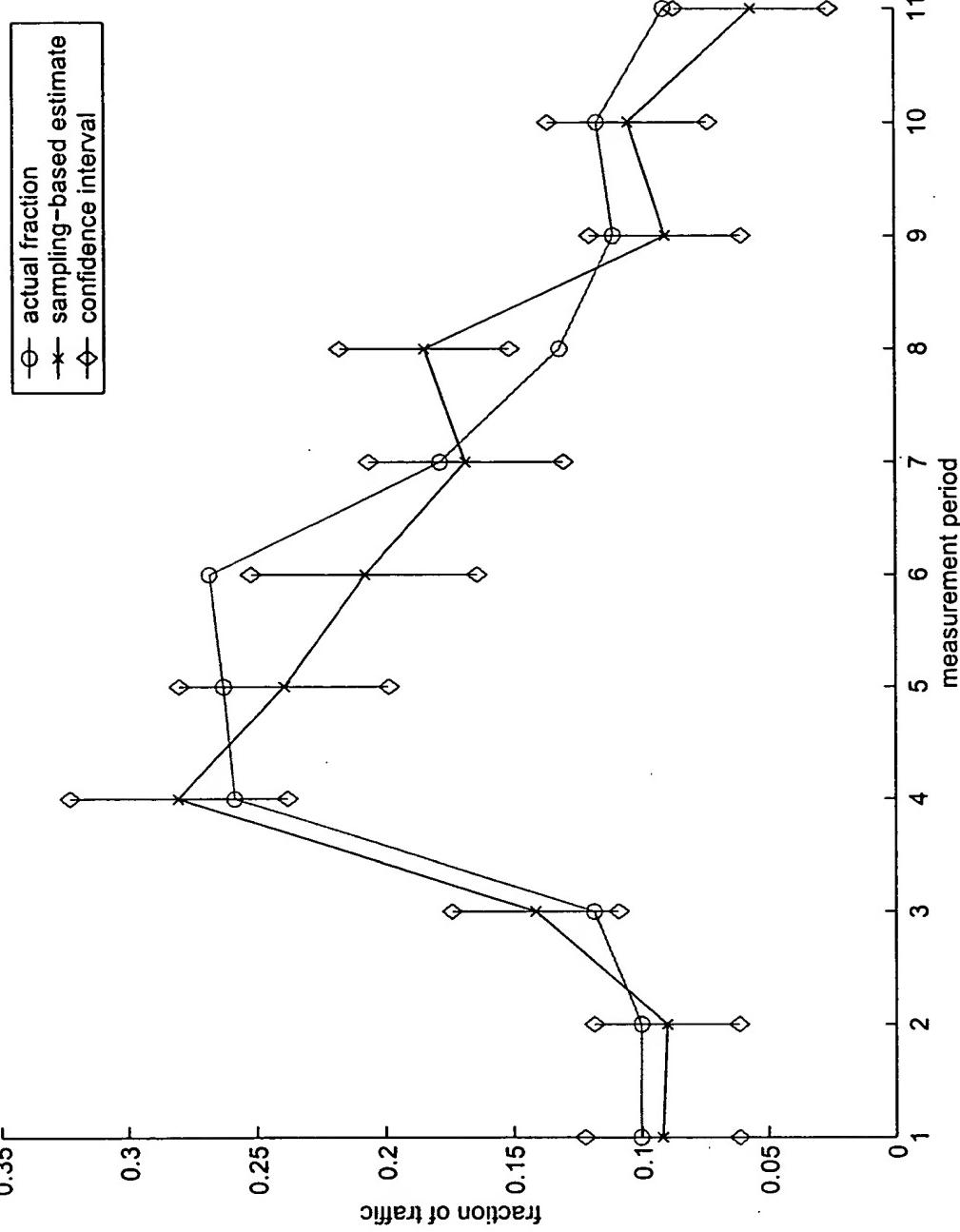
- Experiment: inference from trajectory samples
 - Estimate fraction of traffic from customer
 - Customer traffic: small source address subset



- Fraction of customer traffic on backbone: μ
Estimator: $\hat{\mu} = n_{c,b}/n_b$
 $n_{c,b}$: # unique labels common on both links
 n_b : # unique labels on backbone link
- Ingress link and source address correlated

Estimated Customer Traffic ($c = 10^3$ [bits/epoch])

Real and estimated fraction of packets with specific source prefix ($c=1000$ bit)



Estimated Customer Traffic ($c = 10^4$ [bits/epoch])

Real and estimated fraction of packets with specific source prefix ($c=10000$ bit)

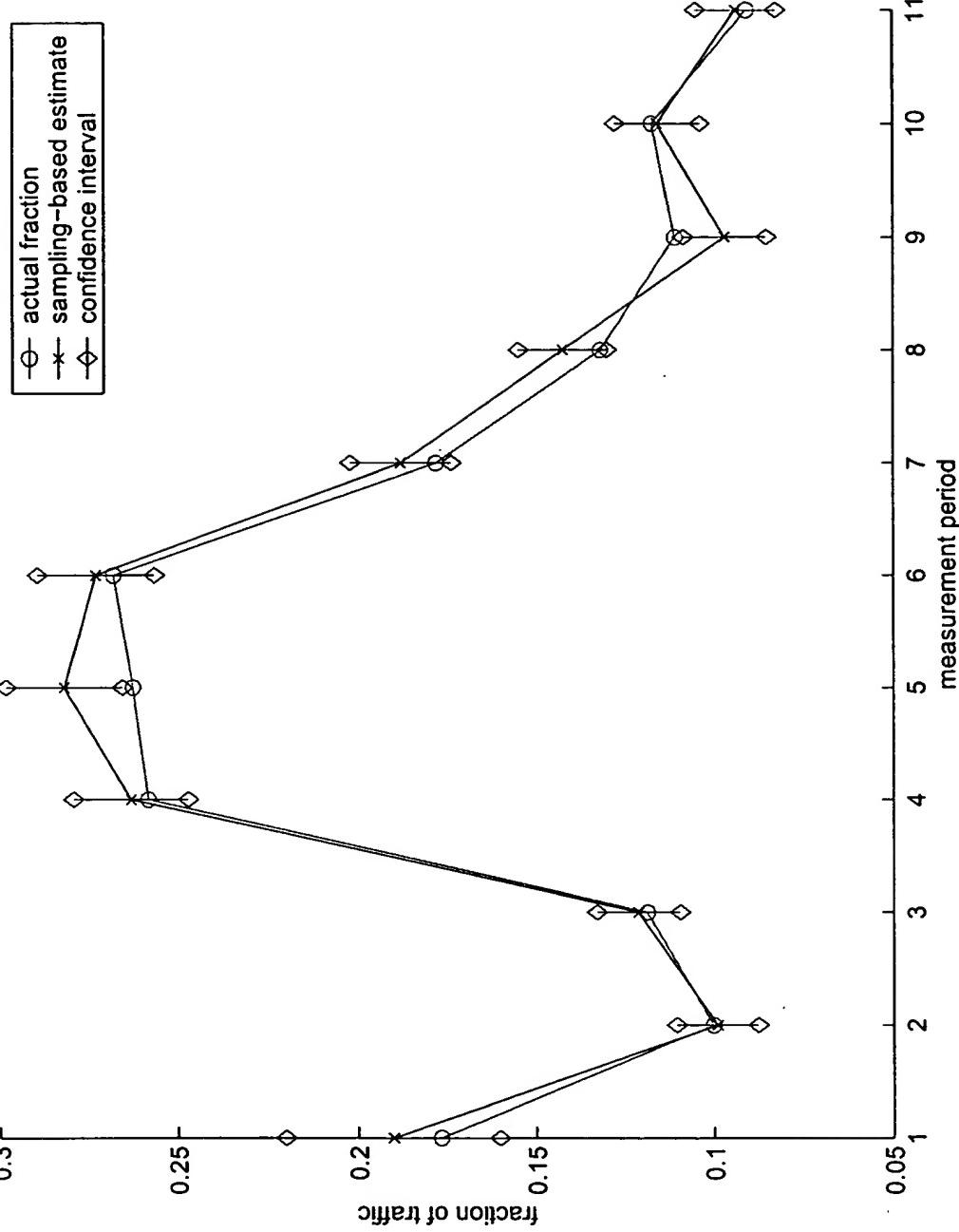


FIGURE 11

Sampling Device Implementation

